

Context Awareness and Its Applications

Masaki Matsudaira Kyoko Hoshikawa
Kohei Taki

A US-PALA™,¹⁾ engine is being developed to facilitate services that can respond to “individual” situations (context-aware) using the behavior logs of a diverse range of “individuals”, featuring “non-discriminatory individualization”, through endeavors by OKI to realize ubiquitous services. “US-PALA” is an abbreviation for the Ubiquitous Service by Personal Activity Log Analysis and it is a means to realize context-aware services through the analyzation of behavior logs as well as by means of behavior predictions and a context estimating technology. The concept of the US-PALA engine is shown in Fig. 1.

With the US-PALA engine it is feasible for a travel agency web site to offer individualized tours by way of selecting tourist attractions that may be of interest to a particular individual, based on his or her past purchases of products and web pages viewed in the past. It is also potentially possible to build services that collect information in advance and provide information considered to be necessary for operating a personal computer (behavior), based on the positional information of an employee, the operating history of the personal computer and schedules.

This paper introduces an outline of the US-PALA engine, technology and service prototypes utilizing the US-PALA engine.

Behavior log and context

Let us first ascertain the meaning of “behavior log” and “context”.

A behavior log is an archive of the previous activities of an individual leading up to the present time. Such a log includes a transfer or relocation history, purchase history, e-mail archive, personal computer operating history, television viewing history, household appliance usage history or biometric information, such as heart rate or blood pressure. A mixture of diversely varying data formats is available for behavior logs, in terms of consecutive values, dispersion values, range and size of values, as well as the frequency of data updates. Furthermore, the means used to obtain such data also varies greatly, from sensor devices to information devices, such as personal computers and mobile phones or household appliances, host computer systems or web services. How such a diverse range of data is integrated and processed is a critical issue.

Context, on the other hand, has been defined in the following manner by Dey and his associates at the Georgia Institute of Technology^{*1)}:

Service user (personal computers, mobile phones, home servers)

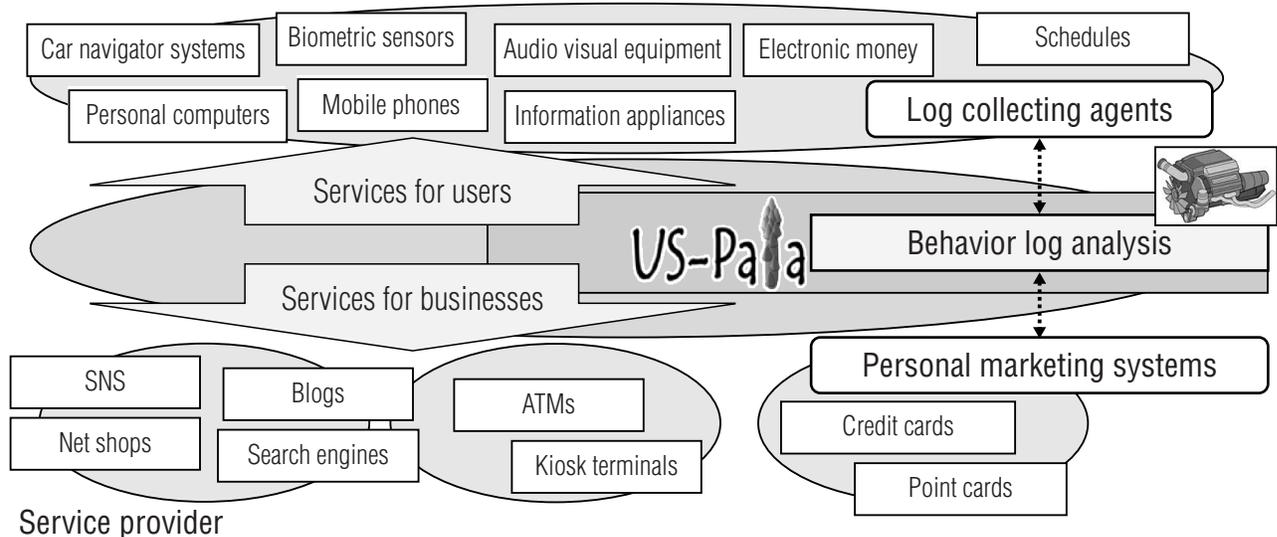


Fig. 1 Concept of US-PALA engine

*1) US-PALA is a trademark of Oki Electric Industry Co., Ltd.

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

While keeping in mind the definition provided by Dey and his associates, we consider context to be situations of an individual or an environment from which behavior log data can be obtained. Context, therefore, is given meaning based on log data. Behavioral situations, such as “standing still”, “walking”, “riding a train”, as well as individual steps for work operations or preferences of an “individual” for example, correspond to such context.

Dey and his associates, furthermore, define a context-aware system in the following manner:

A system is context-aware if it uses context to provide relevant information and/or services to the user, where the relevancy depends on the user’s task.

Based on this definition, we consider context-aware services to be services that provide necessary information or functions when they are needed, using the context of an “individual”, rather than simply personalizing “individualization” services.

Outline of US-PALA engine

The US-PALA engine features a function to build behavior models, based on behavior logs using the Bayesian network technology and natural language processing technology, as well as a function for forecasting behaviors and estimating context based on the present conditions. The configuration of the US-PALA engine is shown in **Fig. 2**.

The behavior model building function loads behavior logs of an “individual” that have been collected using mobile devices, agents embedded in personal computers, sensors and sensor networks allocated to spaces, as well as web services that distribute environmental data and builds a stochastic behavior model. A stochastic cause and effect relationship is learned, for example, from the relationship between the departure time from the office and positional information (or purchase history at a book shop). In this manner, a model can be built, such as “when the departure time from the office is earlier than usual, a high probability exists that the person will stop by the book shop”.

The behavior forecast and context estimating function forecast behaviors or estimate context through the entering of observable conditions, such as the position of an “individual”, time or the immediately preceding behavior, based on the behavior model of the “individual”. Activity support services, notification services or recommendation services can be built in accordance with such forecasts or estimations. The Bayesian network and natural language processing technologies are described below.

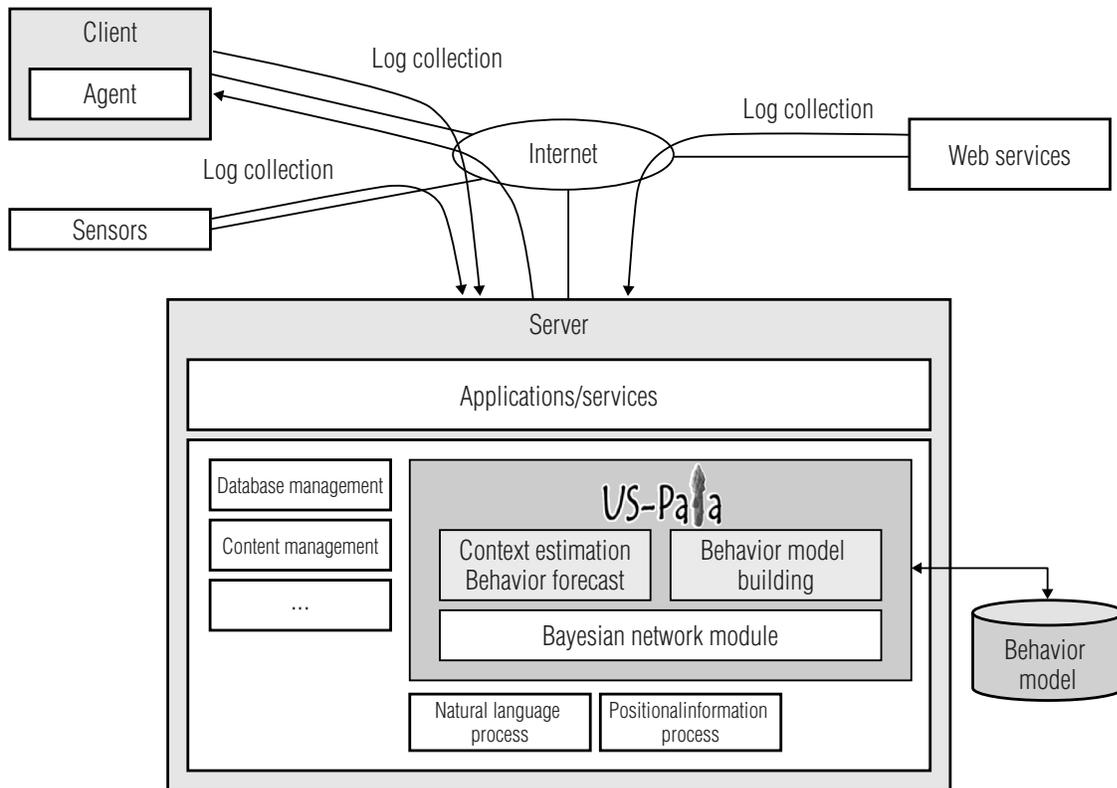


Fig. 2 Overall configuration of US-PALA engine

Bayesian network

A Bayesian network is a technology for efficiently expressing the joint probability distribution of variables through the detection of dependencies between multiple variables^{2), 3)}. A Bayesian network is comprised of a directed graph, which expresses the dependency between variables, as shown in Fig. 3 and a conditional probability table for each individual variable. A node in the directed graph corresponds to a variable. A directed side indicates that a dependency exists between variables that correspond to the nodes it links. The directed side from Node X_1 to Node X_3 indicates that the corresponding variable X_3 is dependent on variable X_1 . A conditional probability table exists for each variable, indicating the variables on which the variable has dependencies and relevant conditions.

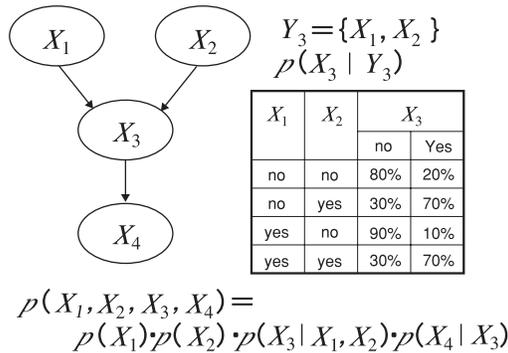


Fig. 3 Example of Bayesian network

With the Bayesian network the approximation of simultaneous probability is derived in the following manner by way of an assumption that absolutely no dependencies exist outside those expressed by the directed graph:

$$p(X_1, \dots, X_n) = \sum_i p(X_i | Y_i)$$

The Y_i however, represents a group of variables to which X_i is dependent.

The US-PALA engine uses this Bayesian network to express the behavior model of an "individual" and forecasts behaviors and context estimations. In other words, variables of the Bayesian network and the conditional probability are learned from a diverse range of behavior logs of an "individual" to forecast behaviors and context estimations using entered observable conditions, such as the position of the "individual", the timing or immediately preceding behavior.

The Bayesian network is able to flexibly express uncertain behaviors of a person using a framework of probability.

Natural language processing and knowledge processing

Another technical characteristic of the US-PALA engine features an integration of natural language processing and knowledge processing, which have been nurtured at OKI over many years.

Preferences of an "individual" in particular are often expressed within the text contained in a behavior log. A person who purchased tickets for a jazz concert, for example, may have a great love for jazz piano or may be a fan of jazz from the 1960s. Such preferences are often expressed in the text found in the blog of an individual or in the introductory phrases of products. It is possible to better express the preferences of an "individual" by means of extracting the keywords from such text.

The Term Frequency/Inverse Document Frequency (TF/IDF) method is a known method used to extract keywords from text. This is based on the concept that terms appearing frequently in a document characterize the document, whereas terms that appear in many documents are not characteristic, the degree of importance $w(t, d)$ of a term t in a document d is defined by the following formula:

$$w(t, d) = tf(t, d) \cdot idf(t)$$

The $tf(d, t)$ here is the frequency with which the term t appears in document d , whereas $idf(t)$ is derived from a number of documents $df(t)$, including the number of documents N and the term t using the following formula:

$$idf(t) = \log \frac{N}{df(t)}$$

The TF/IDF method is effective for relatively long documents, such as a thesis or newspaper article, but modification is needed, since the appearance frequency for terms, tf , often becomes 1 in shorter documents, such as product descriptions. We are in the process of expanding the keywords by way of building a thesaurus that describes the relationship between the terms and ontological dictionary⁴⁾.

It is also necessary to devise a means to process the keywords using the Bayesian network. Simply by entering the keywords as variables in the Bayesian network results in the use of an enormous amount of processing time, since as many variables as the number of keywords must be processed. The amount of processing is being reduced with the use of a thesaurus and ontological dictionary, as well as by way of determining the amount of mutual existent information through statistical processing.

Service prototypes

The following types of “individualization” and context-aware services can be conceived as an applicable range of the US-PALA engine:

- Financial and retail fields: Notifies the subject about sought-after goods and items when they are required and recommendations for stores or products as deemed necessary based on the specific situations.
- Travel and tourism fields: Supports drivers and travelers depending on specific destinations, routes, travel schedules and situations.
- Health care field: Monitors conditions of a patient and notifies a physician as the needs arise.

The recommendation of books is a popular feature of “individualization” services in financial and retail fields. This service recommends that the relevant books be provided in response to the purchase of a book. Due to the constraints of the formula for recommendation “individualization” this service is performed solely based on the purchase history available on a particular web site, other behaviors are not reflected in such recommendations.

The “individualization” service is also urgently needed in travel and tourism fields. The number of individual sight-seeing tours in recent years, indicates through investigation results⁵⁾ that individual sight-seeing tours dominated the market of domestic travel by 77.8% in the year 2005. Social factors, such as the availability of environments to provide searching and booking services 24 hours a day, due to the popularization of the internet, or the fact that diversified judgments of consumer behavior have become the norm, may also be contributing to this demand. The currently available services are scarcely satisfactory for the needs of users and the need for “individualized” services is becoming even greater.

Such market environments have been considered in the development of our services to recommend shops or sight-seeing destinations in response to the specific preferences of an individual or a particular situation. A display example of such a service is shown in **Fig. 4**.



Fig. 4 Display example of service prototype

Future undertakings

We intend to build an “individualization” service by means of a coordinated linkup of the engine with systems created by OKI and other companies. For example, a user can select where he or she wants to go from a list of recommended tourist attractions and restaurant information for the “individual” and through the sharing of the information with a travel assisting service it will then be possible to easily prepare an “individualized” travel itinerary^{6), 7)}.

References

- 1) Dey, A. K. and Abowd, G. D.: “Towards a Better Understanding of Context and Context-Awareness”, *Proceedings of the CHI2000 Workshop on The What, Who, Where, and How of Context-Awareness*, 2000.
- 2) Kimura: Bayesian Network Software, *The Journal of the Japanese Society for Artificial Intelligence*, Volume 17, Number 5; 2002.
- 3) Honmura and Iwasaki: “Bayesian Network Technology: Modeling of Users, Customers and Analogism of Uncertainty”, Tokyo Denki University Press; 2006.
- 4) Matsudaira, Ueda, Onuma and Morita: “Attempt to Build Ontology Based on Analysis of Web Content and Sort Out Information”, A document from the Fourth Meeting of the Semantic Web and Ontology Study Group, The Japanese Society for Artificial Intelligence; 2003.
- 5) Toray Corporate Business Research, Inc: “Current Status and Issues for the Travel Industry - Steady Footing with Leading Roles Being Played by Young to Middle Age Groups - A TBR Investigative Report”; 2006.
- 6) Kawahara, M: “Platform for Realizing Ubiquitous Services”, *OKI Technical Review*, Issue 210, Volume 74, Number 2, pages 4 to 9; 2007.
- 7) Aramaki, Kao, Nakagawa and Enomoto: “Real/Virtual Fusion Travel Assist Service”, *OKI Technical Review*, Issue 210, Volume 74, Number 2, pages 70 to 75; 2007.

Authors

Masaki Matsudaira: Ubiquitous Service Platform Company, Ubiquitous Technology Dept.-2, Development Team-4

Kyoko Hoshikawa: Ubiquitous Service Platform Company, Ubiquitous Technology Dept.-2, Development Team-4

Kohei Taki: Ubiquitous Service Platform Company, Ubiquitous Technology Dept.-2, Development Team-4